

Improving Remote Sensed Data Products Using Bayesian Methodology for the Analysis of Computer Model Output Grant Number NNX07AV69G – First Year Report

Robin D. Morris, USRA-RIACS
Athanasios Kottas, UCSC
Roberto Furfaro, Barry D. Ganapol, UofA

1st August 2008

Abstract

We report on progress made during the first year of this grant. Work in this first year has been in four categories, and will provide the foundation for work in the second year.

The first category has been the extensive revision of a paper submitted to the IEEE Transactions on Geoscience and Remote Sensing. This interaction with the reviewers from the remote sensing community was very fruitful, resulting in a significantly improved paper, and a deeper understanding of some of the key issues within the remote sensing community.

One of these key issues is the inversion of models to produce data products. We have identified data sets which we will use in future work to produce validated inverse estimates from the LCM model.

Another key issue is the use of bi-directional reflection coefficients, as opposed to the hemispherically integrated reflection coefficient produced by the current version of the LCM algorithm. This has spurred two parallel efforts, the further development of the LCM algorithm to enable the computation of bi-directional reflection coefficients, and the study of the sensitivity of the LCM for nadir observations, which can be simulated with the current version of the LCM.

We conclude with plans for year two, and a list of presentations and publications.

1 Revision of paper submitted to IEEE Transactions on Geoscience and Remote Sensing – now accepted for publication

One major work item this year has been the extensive revisions of the paper “A Statistical Framework for the Sensitivity Analysis of Radiative Transfer Models” which has now been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. A final camera-ready version of this paper is attached to this report.

The reviewers made a number of comments on the initial manuscript, which have provided us with valuable and useful feedback from the remote sensing community on our work and how our work fits in with the ways of thinking of the remote sensing community.

The main issue was due to a very strong focus in the community on the production of data products, and the associated model inversion, and hence an assumption that our paper was going

to address model inversion. While that is a goal of the research funded under this grant, the current paper describes preliminary work towards that goal, namely the detailed analysis of the behaviour of the forward model, in terms of the importance of each input to the response of the model. This gives information as to which of the inputs are likely to be able to be accurately estimated when inverting the model. A related issue was that, because of the focus on inversion, the reviewers did not immediately understand the usefulness of a global sensitivity analysis (looking at the properties of the model as all the inputs vary over their valid ranges), rather than a local sensitivity analysis about a fixed point (the fixed point being the result of the inversion for a given data set). Global sensitivity is useful for model analysis, and also because the model will be used for inversion over a wide range of inputs, and so the actual fixed points about which local sensitivity should be computed are unknown.

A second issue was the use of directional-hemispherical reflectances (as produced by the current version of the LCM). The reviewers pointed out that in current practice, bi-directional reflection coefficient are used. As described in section 4 below, we are now developing further the LCM model to allow for the computation of bi-directional reflection coefficients.

The reviewers made a further suggestion that by coupling the LCM to an atmosphere model (such as MODTRAN), and performing the sensitivity analysis at a very large set of closely spaced wavelengths, it would be possible to determine the spectral regions where the detection of leaf biochemistry is most successful, and where atmospheric water vapor does not interfere with leaf water. We regard this as a very interesting suggestion, and will address it in future work.

2 Identification of satellite data corresponding to a particular BIGFOOT site

As discussed above, the primary goal of this research project is to develop methodology for the estimation of biospherical parameters, particularly LAI, that are well-characterized in terms of the uncertainties on these estimates. It is important, therefore, to identify data sets that will allow us to both perform the model inversion to generate LAI estimates with their uncertainties, and also to identify data sets where corresponding ground-truth data is available.

We have identified the Harvard Forest BIGFOOT site as a suitable test region. The Harvard Forest is located at 42.37N and 72.25W. It is a mixed temperate forest. Detailed field data for a 7km \times 7km region around the flux tower site is available from <http://daac.ornl.gov>. We have also identified the most suitable MODIS data set for use for our model inversion research. This is the MOD09 data set, which are atmospherically corrected surface reflectances.

The data considered are Level 3 products in which every pixel is geolocated and arranged in non-overlapping tiles. For this specific exercise, we considered MOD09A1 data which provides MODIS band 1-7 surface reflectance at 500m resolution. It is a Level 3 product and it is the best possible L2G observation over the course of a 8-day period as selected on the basis of high observation coverage, low view angle, absence of clouds or cloud shadow, and aerosol loading.

The MODIS product extracted from the database and containing the Harvard forest as subset has the following ID: MOD09A1.A2008185.h12v04.005.2008195045601.hdf

The MATLAB HDF-EOS tool has been used to extract the surface reflectance from the downloaded data. Figure 1 shows the RGB composite image of the surface reflectance.

In figure 2 we show the section of the data which lies within 1 degree of latitude and longitude of the location of the Harvard Forest. We have also extracted a subset of 10 \times 19 pixels which are



Figure 1: RGB composite image from the selected MODIS data. (band 1→red, band 2→green, band 3 → blue)



Figure 2: Subsection of the MODIS data corresponding to a 1 degree by 1 degree area around the Harvard Forest.

given by the biome map to be broadleaf conifers.

3 Study of the “main effects” and “sensitivity indices” of the LCM for sun angles corresponding to the data identified above, and nadir observations

As discussed, when inverting a radiative transfer model using MODIS data to produce an estimate of LAI, the MODIS data typically provides bi-directional reflection distribution coefficients. In preparation for the completion of the extensions to the LCM to allow the computation of these coefficients (see section 4), we performed a sensitivity analysis of the current version of the LCM for the special case of a BRDF that it can compute – the case of nadir observations, which are a common observation mode for MODIS.

The LCM was run using the illumination angles corresponding to the Harvard Forest data discussed above, for the seven MODIS bands included in the data set. (Note that this is one band less than the case considered in the IEEE paper; the 748nm band is not included in the MODIS data discussed above.) Figure 3 shows the main effects for the LCM in each of these seven bands, and table 1 gives the sensitivity indices.

Comparing these with figure 2 and table III in the IEEE paper (see appendix A) we see that the main effects are for the most part indistinguishable, there being some slight differences in band 3 for LAI and Lignin. The sensitivity indices are also very similar between the two cases, with small differences in band 6, where the LCM is now more sensitive to chlorophyll and less sensitive to LAI.

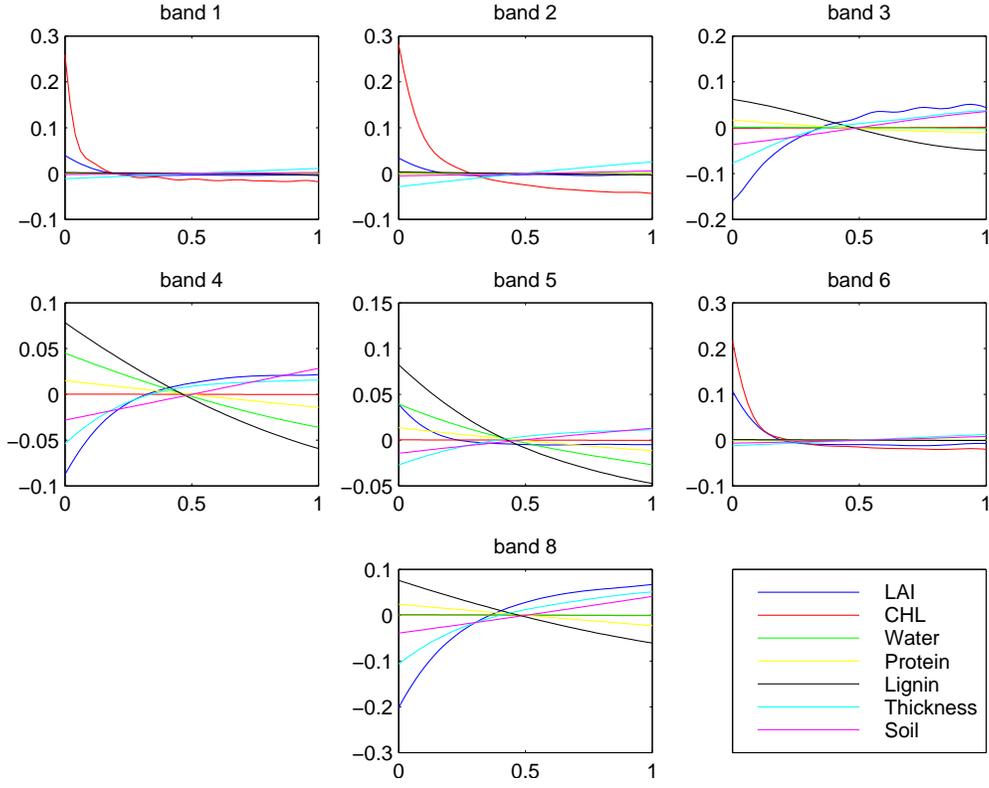


Figure 3: Main effects for the LCM for nadir reflection for the 7 spectral bands given in table 1

	band; wavelength (<i>nm</i>)							
input	1	2	3	4	5	6	7	8
LAI	0.05	0.01	0.43	0.17	0.04	0.23		0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.62		0.00
Water	0.00	0.00	0.00	0.12	0.14	0.00		0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00		0.02
Lignin	0.00	0.00	0.18	0.36	0.54	0.00		0.16
Thickness	0.02	0.05	0.14	0.07	0.05	0.02		0.18
Soil	0.00	0.00	0.07	0.06	0.03	0.01		0.05
Total	0.88	0.90	0.83	0.80	0.81	0.88		0.90

Table 1: Sensitivity indices for the LCM for nadir reflection for the 7 spectral bands

4 Developments of the LCM model to enable the computation of Bi-directional Reflection Distribution Functions

As discussed above, inversion of MODIS data for the production of data products (e.g. LAI) is typically performed using a model that allows for the prediction of bi-directional reflection coefficients. This allows for the more effective use of the data, which is recorded from different observation angles. To this end, we are developing an extension to the LCM radiative transfer model that computes the two-angle reflection coefficients.

At the canopy level, the equation describing the radiative regime inside the vegetation is derived by applying the conservation of photons to the differential element in the phase space. The canopy is assumed to be a “turbid” medium with leaves as scattering elements having a defined normal probability distribution function. The leaf normal orientation influences the scattering type which cannot be considered isotropic. In the current version, one spatial coordinate (canopy depth) and one angle (inclination) are defined as independent variables. We improve the model by implementing a two angle formulation. The two-angle CANMOD equation can be written as following:

$$\left[\mu \frac{\partial}{\partial \tau} + G(\mu) \right] I(\tau, \Omega) = \frac{1}{\pi} \int_{4\pi} d\Omega' \Gamma(\Omega', \Omega) I(\tau, \Omega') \quad (1)$$

where τ is the depth within the canopy in LAI units, and $\Omega = (\theta, \phi)$ is the solid angle. The boundary conditions are

$$\begin{aligned} I(0, \Omega) &= F_L(\Omega), \mu > 0 \\ I(\Delta, \Omega) &= F_R(\Omega) \\ &= \frac{r_s}{\pi} \int_{2\pi} d\Omega' |\mu'| I(\tau, \Omega'), \mu < 0 \end{aligned} \quad (2)$$

The first line in equation 2 describes the condition at the top of the canopy and usually has two contributions coming from direct sunlight and diffuse sunlight. The second line represents the reflected radiance from the bottom surface at $\tau = \Delta$ resulting from a partially Lambertian reflecting soil.

Equation 1 is an integro-differential equation. $I(\tau, \Omega)$ is the radiance which is defined as the energy per unit area per unit time per unit solid angle transported by photons traveling within the canopy. The inputs of this model are the leaf reflectance and transmittance (optical properties coming from LEAFMOD) as well as LAI, LAD, soil reflectance, sun inclination and azimuthal angle.

Equations 1 and 2 can be numerically solved to determine the radiance exiting the top of the canopy. One of the most important quantities for remote sensing application is the Bidirectional Reflectance Distribution Function (BRDF), defined as:

$$R_f(-\mu, \psi) = \frac{I(0, -\mu, \psi)}{F_L(\mu_i, \psi_i)}$$

This quantity is one of the products of MODIS algorithms processing by the sensing devices. It is also important to note that equation 1 can be solved for different wavelengths to compute the spectral response of the canopy.

A variation of the Discrete Ordinate Method commonly employed to numerically solved transport of neutrons in nuclear devices has been devised to compute the radiant intensity Top-Of-the-Canopy (TOC) as function of the viewing geometry. The discrete equations can be written as follows:

Case $\mu > 0$

$$\begin{aligned} I_{n,m}^i &= \left(1 + \frac{G(\mu_n)\Delta\tau}{2|\mu_n|}\right)^{-1} \left(I_{n,m}^{i-1/2} + \frac{\Delta\tau}{2|\mu_{n,m}|}Q_{n,m}^i\right) \\ I_{n,m}^{i+1/2} &= 2I_{n,m}^i - I_{n,m}^{i-1/2} \end{aligned} \quad (3)$$

Case $\mu < 0$

$$\begin{aligned} I_{n,m}^i &= \left(1 + \frac{G(\mu_n)\Delta\tau}{2|\mu_n|}\right)^{-1} \left(I_{n,m}^{i+1/2} + \frac{\Delta\tau}{2|\mu_{n,m}|}Q_{n,m}^i\right) \\ I_{n,m}^{i-1/2} &= 2I_{n,m}^i - I_{n,m}^{i+1/2} \end{aligned} \quad (4)$$

Here

$$Q(\tau, \Omega_{ij}) = \frac{1}{\pi} \sum_{m=1}^M w_m \sum_{n=1}^N \bar{w}_n \Gamma(\Omega_{nm}, \Omega_{ij}) I(\tau, \Omega_{ij}) \quad (5)$$

$Q(\tau_i, \Omega_{ij}) = Q_{n,m}^i$ is the inscattering source which has been discretized using the Gauss-Legendre integration scheme for the inclination angle and the Gauss-Chebyshev integration scheme for the azimuthal angle.

Equations 3 to 5 are the discrete equations for CANMOD two-angle formulation. They are solved according to the conventional “sweeping technique”. We are in the process of completing the implementation of the numerical scheme in MATLAB. Evaluation of the computational speed is a critical parameter for selecting the appropriate language. MATLAB is initially used for rapid development and testing. Fortran 95 or C/C++ will be employed for improving the computational speed.

5 Progress on computing the full distribution of the sensitivity indices using Bayesian methods

The approach to global sensitivity analysis developed in the IEEE paper (discussed in Section 1) is based on a Gaussian process (GP) approximation to the LCM output $y = f(\mathbf{v})$, where $\mathbf{v} = (v_1, \dots, v_7)$ is the vector of the 7 LCM inputs considered in the sensitivity analysis, i.e., LAI, chlorophyll, water fraction, protein, lignin/cellulose, thickness, and soil. The LCM approximation was based on maximum likelihood estimates for the parameters of the isotropic GP model, specifically, the mean $E(f(\mathbf{v})) = \mu$, variance $\text{Var}(f(\mathbf{v})) = \sigma^2$, and the scale parameters of the product correlation form

$$\text{Corr}(f(\mathbf{v}), f(\mathbf{v}')) = \exp\left(-\sum_{\ell=1}^7 \frac{(v_\ell - v'_\ell)^2}{\gamma_\ell}\right)$$

where $\mathbf{v}' = (v'_1, \dots, v'_7)$.

Sensitivity analysis focuses on the main effects

$$E(Y | v_j) - E(Y) = \int_{\mathbf{v}_{-j}} f(\mathbf{v}) dH(\mathbf{v}_{-j} | v_j) - E(Y), \quad j = 1, \dots, 7,$$

and the sensitivity indices

$$S_j = \frac{\text{Var}(E(Y | v_j))}{\text{Var}(Y)}, \quad j = 1, \dots, 7.$$

Here, \mathbf{v}_{-j} denotes all the elements of \mathbf{v} excluding v_j , and $H(\mathbf{v})$ is the uncertainty distribution consisting of independent uniform components for each input over specified ranges.

The likelihood-based approach enables relatively straightforward computation of point estimates for the main effects, including a measure of the uncertainty associated with these estimates. However, using point estimates for the GP parameters, $(\mu, \sigma^2, \boldsymbol{\theta} = (\gamma_1, \dots, \gamma_7))$, results in an underestimation of the uncertainty of the main effects. More importantly, the inferential scope for the sensitivity indices is rather limited. Under either a likelihood or empirical Bayes approach, estimation of the S_j is only possible in the form of a ratio of expectations for $\text{Var}(E(Y | v_j))$ and $\text{Var}(Y)$, hence, as a biased point estimate of the expectation of S_j . Indeed, the literature on GP approximation to computer model output appears to be lacking methods for measuring the uncertainty associated with the S_j , which is key for accurate quantification of the relative importance of each input.

We are currently developing fully Bayesian methodology for sensitivity analysis that will address the above issues. In particular, the method will provide full inference for the main effects that incorporates the uncertainty resulting from the use of the GP model as an emulator for the LCM, but also the uncertainty associated with the GP parameters. Moreover, the Bayesian approach will provide proper point estimates for the sensitivity indices, in fact, an entire distribution for the S_j .

The first stage of this work involves obtaining the posterior distribution of the GP parameters through Markov chain Monte Carlo simulation methods. We have designed a Metropolis-Hastings algorithm to obtain posterior samples for $(\mu, \sigma^2, \boldsymbol{\theta})$, and we are in the process of implementing it. The next step involves obtaining posterior realizations for the LCM output function $f(\cdot)$ under the GP approximation (and up to a grid of desired detail over the input space). This will be accomplished using the posterior samples for the GP parameters and the predictive distributions under the GP model. Finally, conditional on each predicted output realization, posterior realizations for the main effects and posterior samples for the sensitivity indices can be obtained using their definition and applying standard numerical or MC integration techniques. By conditioning on the predicted output function (i.e., by working with the posterior realizations for $f(\cdot)$), we can obtain the full posterior distribution of the sensitivity indices.

When complete, this work will be described in a paper to be submitted to an applied statistics journal.

6 Plans for Year Two

Work in year two will build on the foundation outlined above, and will cover the following areas.

- Completion of the extensions to the LCM to allow for the computation of BRDF coefficients.
- Completion of the code to compute the full distribution of the sensitivity indices.

- Investigation of the hierarchical GP model for different LAD values.
- Refinement of the priors used on the LCM inputs.
- Initial work on developing GP priors for the bias function for inversion.
- Initial work on validation of the model using BIGFOOT data.

We will also further consider the Dirichlet Process Mixture Model as an alternative nonparametric model for certain aspects of the analysis. We will also prepare papers describing the extensions to the LCM and model validation.

7 Presentations and Publications

- “Global Sensitivity Analysis of Leaf-Canopy Radiative Transfer Models for Analysis and Quantification of Uncertainties in Remote Sensed Data Product Generation” Presented at the *AGU Fall Meeting*, San Francisco, 2007.
- “A Statistical Framework for the Sensitivity Analysis of Radiative Transfer Models” Accepted for publication in *IEEE Transactions on Geoscience and Remote Sensing*, April 2008.

A Paper accepted for publication in *IEEE Transactions on Geoscience and Remote Sensing*

A Statistical Framework for the Sensitivity Analysis of Radiative Transfer Models

Robin D. Morris, *Member, IEEE*, Athanasios Kottas, Matt Taddy, Roberto Furfaro and Barry D. Ganapol

Abstract— Process models are widely used tools, both for studying fundamental processes themselves, and as elements of larger system studies. A Radiative Transfer Model (RTM) simulates the interaction of light with a medium. We are interested in RTMs that model light reflected from a vegetated region. Such an RTM takes as input various biospheric and illumination parameters and computes the upwelling radiation at the top of the canopy. The question we address is: which of the inputs to the RTM has the greatest impact on the computed observation?

We study the Leaf Canopy Model (LCM) RTM, which was designed to study the feasibility of observing leaf chemistry remotely. It's inputs are leaf chemistry variables (chlorophyll, water, lignin, cellulose) and canopy structural parameters (LAI, LAD, soil reflectance, sun angle). We present a statistical approach to sensitivity analysis of RTMs, to answer the question posed above. The focus is on global sensitivity analysis, studying how the RTM output changes as the inputs vary continuously according to a probability distribution over the input space. The influence of each input variable is captured through the “main effects” and “sensitivity indices”. Direct computation requires extensive, computationally expensive, runs of the RTM. We develop a Gaussian Process approximation to the RTM output to enable efficient computation. We illustrate how the approach can effectively determine the inputs that are vital for accurate prediction. The methods are applied to the LCM with 7 inputs and output obtained at 8 wavelengths associated with MODIS bands that are sensitive to vegetation.

Index Terms—radiative transfer model, MODIS, sensitivity analysis, main effects, sensitivity index, Gaussian Process

I. INTRODUCTION

The accurate estimation of properties of the biosphere is critical for our understanding of the Earth's coupled system. The atmosphere, oceans and land comprise a complex, coupled dynamical system, and valid statistical prediction of the properties of this system, and its changes, require inputs that are both accurate and have their uncertainties accurately quantified.

The study of the biosphere is dependent on models – mathematical abstractions of the systems themselves, which are sufficiently simplified to allow for mathematical or computational analysis in reasonable amounts of time. Study of these models can give important information about the systems being modeled, but shortcomings in the models, where they

differ from reality, must not be overlooked. Models will have limitations due to the modeling philosophy chosen – the set of simplifying assumptions used by the scientist. Indeed, modeling uncertainty comes from a combination of ignorance of natural variability and the impossibility of precisely modeling the physical phenomena being studied.

The behavior of these models with respect to their inputs is the subject of this paper. Analyzing the uncertainty characteristics of a model is a crucial first step in the use of the model for prediction and inversion. It gives information about the influence of the inputs, both individually and in groups, on the model output, and can give information as to the potential of successfully performing model inversion.

In many cases the model inputs are not easily observable. Instead, the model outputs are measured, and the model inputs must be inferred. Global models require global inputs, and the only effective method for making routine global measurements is via sensors mounted on orbiting satellites. Passive visible/near infra-red sensors measure upwelling radiation, and it is from these measurements that the biospherical parameters of interest must be inferred.

This inference process is complex. It is the inversion of the process of sunlight passing through the atmosphere, being reflected off vegetation on the ground, and then passing again through the atmosphere before being detected by the satellite mounted sensor. The dominant sources of uncertainty in this scenario are the uncertain process models that enter the estimation. The uncertainty due to the process models will almost certainly be much larger than the uncertainty due to noise in the sensor [1]. MODIS [2] has an SNR of between 74 and 910 in the near infra-red bands [3] and is radiometrically very well calibrated in other bands [4], [5]. In the brief outline above we have two models, one for the propagation of light through the atmosphere, and the second for the reflection of light by the vegetation on the ground. It is the uncertainty characteristics of the second of these process models that we will analyze in this paper.

Analyzing, quantifying and reporting the uncertainty in remote sensed data products is of great importance. It is the only way in which the uncertainty of further analyses using these data products as inputs can be quantified. Analyzing the source of the data product uncertainties can identify where the models must be improved, or where better input information must be obtained. Both of these aspects are known; the editorial for the Special Issue on Global Land Product Validation [6] wrote

users need access to quantitative information on product uncertainties

R. D. Morris is with USRA-RIACS, 444 Castro St, Suite 320, Mountain View, CA 94041. rdm@riacs.edu. Tel (650) 966 5035; fax (650) 966 5021. Corresponding author.

A. Kottas is with the Department of Applied Mathematics and Statistics, University of California, Santa Cruz.

M. Taddy is with the University of Chicago Graduate School of Business.

R. Furfaro and B.D. Ganapol are with the Department of Aerospace and Mechanical Engineering, University of Arizona.

and that

[m]aking quantified accuracy information available to the user can ultimately provide developers the necessary feedback for improving the products.

The work in this paper is intended to contribute to this larger goal. We provide tools to allow model builders to better analyse the characteristics of the models that will subsequently be used for inversion. Developments of the tools we provide here will also allow for the characterisation of the uncertainties in the inverse process.

There has been significant work towards the goals outlined above, but there is still much to be done. For example, the current MODIS LAI/fPAR (Leaf Area Index/fraction of Photosynthetically Active Radiation) algorithm has been improved continuously since the satellite's launch. The main improvements have been in the use of a better biome map (reducing the uncertainty in that input); improvements in atmospheric correction; and improved models of surface reflectance from different biomes [7].

These improvements have reduced the uncertainty in the resulting data product, but have not necessarily improved the quantification of the uncertainties, and have not specifically addressed the statistical identification of the sources of the uncertainties. Here we will address one aspect of this overall process. Models of surface reflectance are typically Radiative Transfer Models (RTMs). We analyze in detail the effects of the inputs to an RTM in terms of the sensitivity of the RTM's output to each of the inputs. Specifically we analyze the Leaf Canopy Model (LCM) RTM [8], used as a surrogate for the RTM used as the basis for the MODIS production algorithm [9]. See section II for a discussion of the LCM. In section III, we use and develop methods from the statistical literature on sensitivity analysis [10] to compute the main effects, which graphically show the relative importance of each input on the RTM output, and the sensitivity indices, which give a measure of the expected amount by which the uncertainty in the output would be reduced if the true value of the input was known.

A 1999 paper, [11], discussed the state of sensitivity analysis in the remote sensing and geoscience domains. At that time the analyses were typically very basic, looking only at one variable at a time, and based around a fixed operating point. A number of suggestions as to better methods were made, principally the Fourier Amplitude Sensitivity Test [10]. This suggestion does not seem to have been adopted – the number of papers that cite [11] is small, and the number that adopt the suggestions, smaller still. For example, [12] uses ideas from the design of experiments, but does not compute sensitivity indices. While discussing sensitivity indices, the analysis in [13] is based on local sensitivity computations. In [14] sensitivity indices are computed, but the methods used required large numbers of model runs. In this paper we give explicit, computationally efficient methods for computing the main effects and sensitivity indices, as part of a global sensitivity analysis.

Computing the main effects and sensitivity indices requires the evaluation of multidimensional integrals over the input space of the model. Evaluating RTMs can be computationally expensive, and so standard numerical integration methods (e.g. multidimensional quadrature or Monte Carlo integration)

would be computationally prohibitive in terms of the number of times the RTM would have to be run. Instead, we adopt the approach of approximating the RTM by a Gaussian Process (GP) model [15]–[17], a technique known in the statistical literature as *emulation*. A GP provides a flexible nonparametric function approximation that has found wide application as a replacement for neural networks [15]. Early work involving GP response-surface approximations for the analysis of computer experiments includes [18]–[20]. We refer to [21] for background and further references. The GP model approximation can be constructed using a comparatively small number of carefully chosen RTM evaluations. See section IV. Using the GP approximation instead of the actual RTM will introduce uncertainty into the evaluation of the main effects, and the sensitivity indices, but this can also be quantified [22]. See section V. The GP emulator, being a fully specified statistical model, is amenable to further analysis in ways that a set of sample responses of an RTM, or even the implementation of the RTM as a piece of software, is not. It allows for calibration and validation of the model in a principled statistical manner, and as the likelihood in a statistical treatment of the model inversion. See [23], [24] and section VII. The GP emulator provides a unifying framework for this and other problems.

Finally, in section VI we present the main effects and sensitivity indices for the LCM RTM, and show how they enable the identification of the relative importance of each input to the model output. This also gives information as to how well these inputs can be predicted from observations of the model output at different wavelengths.

II. A COUPLED LEAF-CANOPY RADIATIVE TRANSFER MODEL

Over the past decade, in collaboration with the Ecosystem science and technology branch at NASA Ames, the Vegetation Modeling Transport Group (University of Arizona) has developed a coupled Leaf-Canopy Model (LCM) in order to capture the essential biophysical processes associated with the interaction between light and vegetation [8]. LCM was developed to provide a tool to aid in remote sensing as applied to ecosystem dynamics in support of the TERRA platform and it is specifically used to investigate the feasibility of observing chemistry remotely. The model combines two different radiative transfer models, one at leaf level (LEAFMOD) and one at canopy level (CANMOD) to predict the radiative regime inside the vegetation canopy under consideration.

LEAFMOD [25] is the model that simulates the radiative regime inside the single leaf. From a morphological point of view, the leaf element is an extremely complex and rich object. Any model that attempts to describe each single interaction process for the light moving in such a medium will face this enormous complexity. The strength of the LEAFMOD algorithm is its simplicity through natural averaging. The model relies on the fact that while light is moving in a complicated medium, natural averaging occurs in such way that the simpler assumption of isotropic scattering and uniform absorption seems to capture the transport effects. Moreover, the model has the ability to include chemistry as a key element

dominating the absorption process. Different concentrations of chlorophyll, water, lignin and cellulose can be specified to model the optical properties of the single leaf species. The model is calibrated over the LOPEX leaf species archive [26], where experimental leaf property data are stored. The calibration occurs in the sense that the optical properties required by the canopy model are retrieved through a procedure that uses the LOPEX archive as input data.

While LEAFMOD has been specifically designed and implemented to be coupled with the canopy model (CANMOD, see below), it can also be used as a stand-alone module to describe the radiative transfer of photons within leaf media as functions of their morphological structure and biochemical signature. In early deterministic models, two-stream models were used to determine the radiative transfer within leaf structures. The so-called Kubelka-Munk (KM) theories [27] treat the leaf as a plane parallel medium, tacitly assume the scattering to be nearly isotropic and assume weak volume absorption within leaves. By modeling the transport of photons as a diffusion process the computed radiance is subject to large errors in optically thin media and/or in highly absorbing regions [28]. To overcome some of the difficulties associated with KM theories, PROSPECT [29] was established. Within the PROSPECT framework, the leaf is assumed to be modeled as a sequence of transparent plates, each assumed to be rough Lambertian reflectors. Each plate defines the optical properties of the interior of the leaves. Scattering is described by a spectral index of refraction and a parameter describing the leaf mesophyll structure. The absorption coefficients for leaf water and pigments are generally fitted using experimental data, i.e. leaf reflectance and transmittance. By contrast, LEAFMOD relies on rigorous first principles, i.e. the balance of photons. LEAFMOD's advantage stems from the fact that the overall leaf biochemistry can be easily specified and the scattering coefficient calibrated via experimental data and direct model inversion [30].

The CANMOD (CANopy Model) algorithm [8], [31] takes the information coming from LEAFMOD regarding the single leaf characteristic (transmittance and reflectance) and together with canopy structural parameters, (LAI and Leaf Angle Distribution), soil reflectance and sun angle inclination, computes, at any given wavelength, the radiative regime within and at the top of the canopy by solving a radiative transfer equation. The strengths of the model are simplicity and the ability to take into account leaf chemistry, which is important to properly describe the light absorption environment.

Figure 1 shows a flowchart that demonstrates the operation of the coupled algorithm. The algorithm can be explained as follows. The first module uses LEAFMOD in the forward and inverse mode to compute the leaf optical properties (i.e. leaf reflectance and transmittance). The second module uses the CANMOD forward mode to compute the spectral canopy hemispherical reflectance factor. The code requires the specification of the input parameters. In addition to the parameters listed in table I, the model also takes as input wavelength (between 400nm and 2100nm), canopy architecture (LAD - leaf angle distribution) and the sun angle. CANMOD is able to handle four discrete typologies of leaf angle distribu-

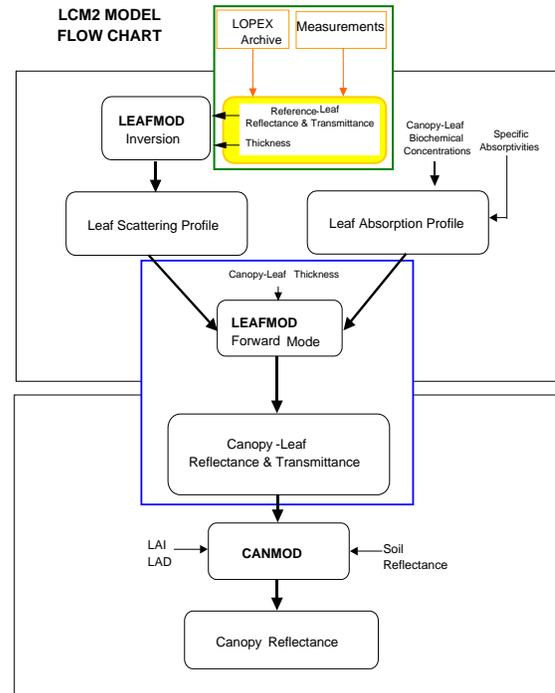


Fig. 1. LCM Flow Chart

input	min	max
LAI	0	8
chlorophyll ($\mu g/cm^2$)	0	100
water fraction	0.1	0.8
protein (g/cm^2)	0.0001	0.001
lignin/cellulose (g/cm^2)	0.0001	0.006
thickness (cm)	0.001	0.01
soil	0.3	1.3

TABLE I

RANGES OF VALUES OF THE INPUTS TO THE LCM. LAI AND WATER FRACTION ARE DIMENSIONLESS. THE SOIL PARAMETER IS MULTIPLIED BY A STANDARD SOIL SPECTRUM. SEE TEXT.

tion, namely planophile (leaves mainly horizontal), erectophile (leaves mainly vertical), plagiophile (leaves mainly at 45 degrees), and extremophile (leaves mainly both horizontal and vertical). LAD is determined largely by knowledge of the biome. Its inference from observational data is difficult [32].

Note that the soil reflectance depends on the wavelength. Indeed, usually the spectral soil reflectance is specified depending on the type of soil of interest. We assumed a typical visible/near infra-red spectrum for a dry soil, and we considered a multiplicative brightness parameter varying between 0.3-1.3 (see table I) to account for the possible variations of the background (soil) reflectance level [33]. This multiplicative parameter is assumed to be wavelength independent.

Once the leaf type is specified, the LOPEX database contains the measured leaf optical properties for the leaf of interest. Nevertheless, we can tune the canopy by considering leaves that are of the same type but with different biochemistry and thickness. This gives the code great flexibility in modeling the effect of biochemistry on the overall canopy reflectance. The algorithm begins by analyzing the leaf under consider-

ation. Assume, for example that the canopy of interest is a maple canopy. The LOPEX database is accessed to retrieve the measured spectral reflectance and transmittance for a nominal maple leaf. Note that, as before, since the wavelength is set, reflectance and transmittance for the nominal leaf are selected for the specific wavelength of interest. The LEAFMOD inverse mode accepts the reflectance and transmittance and retrieves scattering and absorption coefficients. It is assumed that, to first order, the scattering depends on the anatomical structure of the leaf, while the absorption depends only on the biochemical components [25]. Thus, the scattering coefficient for maple leaves is assumed to be the same and it is retained. A new maple leaf having the biochemical components and thickness specified by the inputs is constructed retaining the same scattering coefficient and constructing the new absorption coefficient for the wavelength of interest. Both absorption and scattering coefficient are fed to the LEAFMOD forward mode to compute the reflectance and transmittance of the desired leaf, i.e., the leaf with thickness, water, chlorophyll, lignin and protein specified by the inputs. Reflectance and transmittance are fed to the second module together with LAI, LAD, soil reflectance and sun angle to compute the hemispherical reflectance.

III. SENSITIVITY ANALYSIS

Sensitivity analysis aims to determine how the variation in the output of a model can be apportioned amongst the inputs [21, ch 7]. That is, it attempts to determine how much of the variation seen in the output is due to variation in each of the inputs. The type of sensitivity analysis we are interested in here is *global* statistical sensitivity analysis, looking at how the output changes as all the inputs vary continuously, rather than the more common *local* derivative-based sensitivity analyses, which look at how the output changes as the inputs are each varied about a fixed point [34]. Clearly this latter type of analysis will give limited information about how the output varies for substantial changes in the inputs.

How the inputs vary is determined by a probability distribution that defines the expected distributions of the inputs. Using \mathbf{v} to denote the vector of model inputs, this distribution is $H(\mathbf{v})$. The actual form of this distribution is problem dependent, and dependent on the amount of knowledge available about each input variable. It may be that for some inputs all that can be given is a physically plausible range (e.g. water fraction is limited to the range 0-1), whereas for others a more precise distribution may be known (e.g. the distribution of leaf thickness for a particular tree type may be known from field measurements). The distribution $H(\mathbf{v})$ also encodes correlations between variables that are known to vary together. The authors in [33] give truncated Gaussian distributions for the variables in table I. In this work we use the simpler formulation of independent uniform distributions over the ranges given in table I for each input variable.

A. Main Effects

Denote the response of the model to input \mathbf{v} as $y = f(\mathbf{v})$. The function $f(\mathbf{v})$ can be decomposed as

$$y = f(\mathbf{v}) = \text{E}(Y) + \sum_{i=1}^d z_i(v_i) + \sum_{i<j} z_{i,j}(v_i, v_j) + \dots + z_{1,2,\dots,d}(v_1, v_2, \dots, v_d) \quad (1)$$

where $\mathbf{v} = (v_1, \dots, v_d)$ is d -dimensional (with $d = 7$ in our sensitivity analysis of the LCM). The first term is the expected value of $f(\mathbf{v})$, i.e.,

$$\text{E}(Y) = \int_{v_j, j=1\dots d} f(\mathbf{v}) dH(\mathbf{v})$$

and the next d terms are the *main effects*, given by

$$\begin{aligned} z_i(v_i) &= \text{E}(Y|v_i) - \text{E}(Y) \\ &= \int_{\mathbf{v}_{-i}} f(\mathbf{v}) dH(\mathbf{v}_{-i}|v_i) - \text{E}(Y) \end{aligned} \quad (2)$$

where \mathbf{v}_{-i} denotes all the elements of \mathbf{v} except v_i . The later terms of the decomposition are the interactions. They give information about the combined influence of two or more inputs taken together. We will not consider them further here.

Plotting the main effects, $z_i(v_i)$ for each i gives a visual impression of the relative importance of each input to the variation in the output. This visual impression is heightened if the inputs are normalized (to the range 0-1, for example, for uniformly distributed inputs), allowing all the main effects to be plotted together on the same plot. See section VI where we present main effects plots for the LCM RTM.

To compute the main effects requires the evaluation of a $(d - 1)$ -dimensional integral. For even moderately complex functions $f(\mathbf{v})$ it will be impossible to evaluate this integral analytically; indeed, for most cases of interest an analytic form for $f(\mathbf{v})$ does not exist, rather, $f(\mathbf{v})$ only exists as a computer program. In these cases the $z_i(v_i)$ must be computed numerically. If evaluating $f(\mathbf{v})$ for a given \mathbf{v} requires appreciable computation then the standard methods of numerical integration, multidimensional quadrature and Monte Carlo integration, will be too computationally intensive to be practical. It is therefore useful to approximate $f(\mathbf{v})$ in such a way that the integrals required can be evaluated analytically. This allows the straightforward computation of the main effects of the approximation, and also the computation of the uncertainty introduced by the approximation to $f(\mathbf{v})$. This is given in sections IV and V, in particular, details of the Gaussian Process approximation we use for $f(\mathbf{v})$ are provided in section IV, and its application to computing the main effects and sensitivity indices is developed in section V.

B. Sensitivity Indices

The sensitivity indices are based on the variances of the terms in the decomposition of $f(\mathbf{v})$ given in equation 1. Specifically, consider

$$V_i = \text{Var}\{\text{E}(Y | v_i)\} = \text{E}[(\text{E}(Y | v_i))^2] - (\text{E}(Y))^2.$$

This is the expected amount by which the uncertainty in y will be reduced if we learn the true value of v_i [22]. It thus gives a measure of how much of the variance of y is due to input v_i . The V_i 's can be normalized to

$$S_i = V_i / \text{Var}(Y)$$

so that the sum of all the S_i 's and higher-order terms ($S_{i,j}$, $S_{i,j,k}$, etc.) is unity. Thus the value of S_i gives the relative importance of input v_i . The S_i 's can also be used to direct improvements – reducing the uncertainty on the input with the largest S_i will have the greatest effect in reducing the uncertainty of the model output. This can be used to direct data collection work.

Computing the V_i 's and S_i 's can be complex, even under the Gaussian Process approximation to $f(\mathbf{v})$. See section V for details.

IV. APPROXIMATING THE LCM USING A GAUSSIAN PROCESS

As discussed in the previous section, computing the main effects and sensitivity indices requires evaluating multidimensional integrals over arguments that include the RTM response $f(\mathbf{v})$. There are two approaches available, either evaluating a numerical approximation of the integral itself, or forming an approximation to the argument of the integral, where the approximation enables the integrals to be evaluated analytically. The choice between these two approaches depends on a number of factors. In terms of the required computation, the trade-off is between the numerical evaluation of the integral (typically via Monte Carlo integration), and the computation required to estimate the parameters of the approximation.

Regarding the main effects, direct numerical approximation of the integrals in equation 2 requires evaluations of $f(\mathbf{v})$ over a sufficiently dense grid in \mathbf{v} . It will thus typically be feasible under small to moderate dimensions for the input space and for computationally reasonable functions $f(\mathbf{v})$ (which is indeed the case for the LCM). However, even for computationally inexpensive models $f(\mathbf{v})$, the same approach for the sensitivity indices V_i becomes substantially more challenging to implement for moderate number of inputs and is, arguably, not viable for high-dimensional input spaces. This becomes clear by inspection of the integrals required for the evaluation of the $E[(E(Y | v_i))^2]$; see Appendix I and II. Hence, in general, for global sensitivity analysis there is clear utility in approximating the model function $f(\mathbf{v})$ even for computer models that are relatively inexpensive to evaluate. As a concrete example, in [35] the analysis of a 10-dimensional fire-propagation model required 10^6 simulations to perform global sensitivity analysis. By contrast, a GP emulator can be built using much fewer samples – only 250 were needed to emulate the LCM (see section VI).

As importantly, looking beyond sensitivity analysis, the construction of a statistical model as an emulator for the model output $f(\mathbf{v})$ provides scope for different types of practically important probabilistic analyses of the computer model. This has been discussed briefly in section I and is elaborated in section VII.

The approximation that we use for the LCM is provided by a Gaussian Process. Gaussian Processes (GPs) are probability distributions over *functions*. Rather than placing a distribution over a (small) set of parameters, a GP places a distribution directly over the function of interest. Under a GP probability model for function $f(\cdot)$, the joint distribution of $(f(\mathbf{v}_1), \dots, f(\mathbf{v}_k))$ is multivariate Gaussian, for any finite set of input points $\mathbf{v}_1, \dots, \mathbf{v}_k$. It is this property that allows for tractable computation – whilst the GP is defined over an infinite dimensional quantity (the continuous function $f(\mathbf{v})$), any computation is necessarily done over only a finite set of locations.

A GP is specified by its mean function, $E(f(\mathbf{v}))$, and its covariance function, $\text{Cov}(f(\mathbf{v}), f(\mathbf{v}'))$. The flexibility of choosing and adapting the mean and covariance functions allows a GP model to be successfully used to approximate a wide spectrum of functions $f(\mathbf{v})$, based on a set of training examples, $\mathbf{d} = \{\mathbf{y}, \mathbf{x}_1, \dots, \mathbf{x}_n\}$, where $\mathbf{y} = (y_1, \dots, y_n)$ and y_i is the response $f(\mathbf{x}_i)$ at observed input point \mathbf{x}_i , $i = 1, \dots, n$. The set of training examples is chosen carefully to optimally sample the input space. Here, we used a Latin Hypercube design [36] to choose the set of inputs to the LCM. The other choices made were to use a constant mean function, $E(f(\mathbf{v})) = \mu$, a constant variance $\text{Var}(f(\mathbf{v})) = \sigma^2$, and the product Gaussian correlation function

$$\text{Corr}(f(\mathbf{v}), f(\mathbf{v}'); \boldsymbol{\theta}) = \exp\left(-\sum_{\ell=1}^d \frac{(v_\ell - v'_\ell)^2}{\gamma_\ell}\right)$$

where $\boldsymbol{\theta} = (\gamma_1, \dots, \gamma_d)$, and d is the dimension of the input space. The γ parameters give a measure of the scale over which the function $f(\mathbf{v})$ varies in each input dimension, and σ^2 , the variance of the GP, determines the overall scale of $f(\mathbf{v})$. Using this mean and correlation function, the GP defines the joint distribution

$$p(\mathbf{y} | \boldsymbol{\theta}, \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2} |C(\boldsymbol{\theta})|^{1/2}} \times \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mu \mathbf{1}_n)^T C^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mu \mathbf{1}_n)\right) \quad (3)$$

where $C(\boldsymbol{\theta})$ is the correlation matrix with (i, j) -th element $\text{Corr}(f(\mathbf{x}_i), f(\mathbf{x}_j); \boldsymbol{\theta})$, and $\mathbf{1}_n$ denotes an n -dimensional vector with all elements equal to 1.

We use the set of training examples, \mathbf{d} , to estimate the parameters $\{\boldsymbol{\theta}, \mu, \sigma^2\}$ of the GP model using maximum likelihood estimation. From equation 3 the log-likelihood is

$$\mathcal{L} = -\frac{1}{2\sigma^2} (\mathbf{y} - \mu \mathbf{1}_n)^T C^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mu \mathbf{1}_n) - \frac{1}{2} \log |C(\boldsymbol{\theta})| - \frac{n}{2} \log(2\pi\sigma^2). \quad (4)$$

The derivatives of \mathcal{L} with respect to each of the parameters can be straightforwardly derived [16]. Maximizing \mathcal{L} results in a point estimate for the parameters, denoted by $\{\hat{\boldsymbol{\theta}}, \hat{\mu}, \hat{\sigma}^2\}$, that we use when evaluating the main effects. Note that using point estimates for these parameters will cause the uncertainty of the main effects to be underestimated. In future work we will consider a fully inferential Bayesian approach where expectations are also taken with respect to these parameters.

Once the GP model parameters are estimated, the first quantities of interest are the predictive distributions for sets of new inputs, conditioned on the training examples. From the definition of the GP, these distributions will be Gaussian. For a single new input \mathbf{v} the predictive distribution for $f(\mathbf{v})$ has mean

$$m \equiv m(\mathbf{v}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\theta}}, \mathbf{d}) = \hat{\boldsymbol{\mu}} + \mathbf{r}^T(\mathbf{v})C^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}\mathbf{1}_n)$$

and variance

$$S \equiv S(\mathbf{v}; \hat{\boldsymbol{\mu}}, \hat{\sigma}^2, \hat{\boldsymbol{\theta}}, \mathbf{d}) = \hat{\sigma}^2 (1 - \mathbf{r}^T(\mathbf{v})C^{-1}\mathbf{r}(\mathbf{v})).$$

Here, $\mathbf{r}(\mathbf{v})$ is the $n \times 1$ vector with i -th element given by $\text{Corr}(f(\mathbf{v}), f(\mathbf{x}_i)) = \exp(-\sum_{\ell=1}^d (v_\ell - x_{i\ell})^2 / \hat{\gamma}_\ell)$, and $C \equiv C(\hat{\boldsymbol{\theta}})$ is the observed $n \times n$ correlation matrix with (i, j) -th element given by $\exp(-\sum_{\ell=1}^d (x_{i\ell} - x_{j\ell})^2 / \hat{\gamma}_\ell)$. Recall that the \mathbf{x}_i are the input values of the training examples.

The joint predictive distribution for $(f(\mathbf{v}), f(\mathbf{v}'))$ corresponding to generic inputs $\mathbf{v} = (v_1, \dots, v_d)$ and $\mathbf{v}' = (v'_1, \dots, v'_d)$ is bivariate normal with (2×1) mean vector

$$\mathbf{w} = \hat{\boldsymbol{\mu}}\mathbf{1}_2 + R^T(\mathbf{v}, \mathbf{v}')C^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}\mathbf{1}_n) \quad (5)$$

and (2×2) covariance matrix

$$W = \hat{\sigma}^2 (B(\mathbf{v}, \mathbf{v}') - R^T(\mathbf{v}, \mathbf{v}')C^{-1}R(\mathbf{v}, \mathbf{v}')), \quad (6)$$

where $B(\mathbf{v}, \mathbf{v}')$ is the (2×2) observed correlation matrix for $(f(\mathbf{v}), f(\mathbf{v}'))$ with off-diagonal element given by $\exp(-\sum_{\ell=1}^d (v_\ell - v'_\ell)^2 / \hat{\gamma}_\ell)$, and $R(\mathbf{v}, \mathbf{v}')$ is the $(n \times 2)$ matrix with first column elements $\exp(-\sum_{\ell=1}^d (v_\ell - x_{i\ell})^2 / \hat{\gamma}_\ell)$, $i = 1, \dots, n$, and analogously for the second column elements replacing v_ℓ with v'_ℓ .

V. APPROXIMATING THE MAIN EFFECTS AND SENSITIVITY INDICES USING THE GAUSSIAN PROCESS APPROXIMATION TO THE LCM

To compute the main effects requires evaluating $E(Y | v_j)$, for $j = 1, \dots, d$, and $E(Y)$, as indicated in equation 2. However, we recall that we are approximating the function $y = f(\mathbf{v})$ by a GP model, and we must account for this approximation by computing $E^*\{E(Y | v_j)\}$ and $E^*\{E(Y)\}$, where we use $E^*\{\cdot\}$, $\text{Var}^*\{\cdot\}$ and $\text{Cov}^*\{\cdot, \cdot\}$ to indicate expectation, variance and covariance, respectively, with respect to the GP predictive distributions. We give details of these quantities here.

For the global mean, we have

$$E(Y) = \int_{\mathbf{v}} f(\mathbf{v}) \prod_{\ell=1}^d dH_\ell(v_\ell)$$

where $H(\mathbf{v}) = \prod_{\ell=1}^d H_\ell(v_\ell)$ is the input distribution, comprising independent components $H_\ell(v_\ell)$, which are uniform distributions over ranges (a_ℓ, b_ℓ) , $\ell = 1, \dots, d$. Therefore,

$$\begin{aligned} E^*\{E(Y)\} &= \int E(Y) dN(f(\mathbf{v}); m, S) \\ &= \int_{\mathbf{v}} m(\mathbf{v}) \prod_{\ell=1}^d dH_\ell(v_\ell) \\ &= \int_{\mathbf{v}} \{\hat{\boldsymbol{\mu}} + \mathbf{r}^T(\mathbf{v})C^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}\mathbf{1}_n)\} \prod_{\ell=1}^d dH_\ell(v_\ell) \\ &= \hat{\boldsymbol{\mu}} + \mathbf{T}^T C^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}\mathbf{1}_n), \end{aligned} \quad (7)$$

where \mathbf{T} is the $n \times 1$ vector with i -th element given by $\prod_{\ell=1}^d \left\{ \int_{a_\ell}^{b_\ell} \exp(-(v_\ell - x_{i\ell})^2 / \hat{\gamma}_\ell) (b_\ell - a_\ell)^{-1} dv_\ell \right\}$.

Regarding the conditional expectation $E(Y | v_j)$, for each value u_j of the j -th input, we have

$$E(Y | u_j) = \int_{\{v_\ell: \ell \neq j\}} f(v_1, \dots, u_j, \dots, v_d) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell)$$

and thus

$$\begin{aligned} E^*\{E(Y | u_j)\} &= \int E(Y | u_j) dN(f(v_1, \dots, u_j, \dots, v_d); m, S) \\ &= \int_{\{v_\ell: \ell \neq j\}} m(v_1, \dots, u_j, \dots, v_d) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) \\ &= \hat{\boldsymbol{\mu}} + \mathbf{T}_j^T(u_j)C^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}\mathbf{1}_n), \end{aligned} \quad (8)$$

where $\mathbf{T}_j(u_j)$ is the $(n \times 1)$ vector with i -th element given by the following expression

$$\begin{aligned} &\exp\left(-\frac{(u_j - x_{ij})^2}{\hat{\gamma}_j}\right) \times \\ &\prod_{\{\ell: \ell \neq j\}} \left\{ \int_{a_\ell}^{b_\ell} \exp\left(-\frac{(v_\ell - x_{i\ell})^2}{\hat{\gamma}_\ell}\right) \frac{1}{b_\ell - a_\ell} dv_\ell \right\}. \end{aligned} \quad (9)$$

The previous expressions provide point estimates for all main effects associated with the d inputs. In particular, for each input $j = 1, \dots, d$, $E^*\{E(Y | u_j)\}$ can be computed over a grid of u_j values to obtain point estimates for the functions $E(Y | u_j)$ (or for $E(Y | u_j) - E(Y)$ using also $E^*\{E(Y)\}$). These estimates can be compared graphically (linear transformations can be applied so that all inputs are on the same scale).

For a measure of the uncertainty associated with these estimates, we use

$$\text{Var}^*\{E(Y | u_j)\} = E^*\{(E(Y | u_j))^2\} - (E^*\{E(Y | u_j)\})^2.$$

Because we already have the expression for $E^*\{E(Y | u_j)\}$ from the derivation above, what is needed is an expression for $E^*\{(E(Y | u_j))^2\}$. This derivation is given in Appendix I, resulting in

$$\text{Var}^*\{E(Y | u_j)\} = \hat{\sigma}^2 (e - \mathbf{T}_j^T(u_j)C^{-1}\mathbf{T}_j(u_j))$$

where e is given by 15 of Appendix I.

The sensitivity indices are defined by

$$S_j = \frac{\text{Var}(E(Y | u_j))}{\text{Var}(Y)}, \quad j = 1, \dots, d.$$

Computing $E^*\{S_j\}$ cannot be done analytically, even under the GP approximation, so we approximate it by computing the ratio of $E^*\{\text{Var}(E(Y | u_j))\}$ and $E^*\{\text{Var}(Y)\}$. (In future work we will use a Bayesian approach implemented via Markov chain Monte Carlo (MCMC) methods [37] to estimate the entire distribution of S_j under the GP approximation, allowing the uncertainty of the sensitivity indices to also be determined.) We have that

$$E^*\{\text{Var}(E(Y | u_j))\} = E^*\{E[(E(Y | u_j))^2]\} - E^*\{(E(Y))^2\}$$

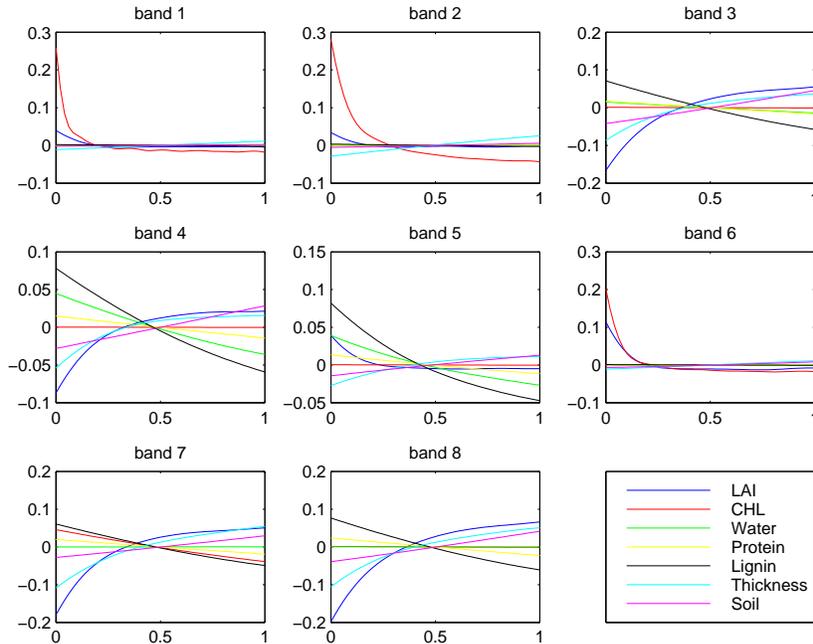


Fig. 2. The main effects for the LCM RTM

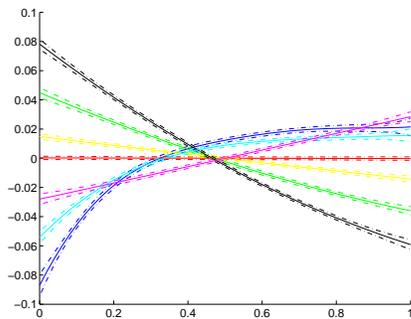


Fig. 3. The uncertainty in the main effects due to using the GP approximation to the LCM RTM. Band 4. Line colours as in figure 2.

and

$$E^*\{\text{Var}(Y)\} = E^*\{E(Y^2)\} - E^*\{(E(Y))^2\}.$$

The expressions for these terms are not difficult to derive, though care is needed. They are given in Appendix II.

VI. RESULTS

The proposed methodology has been applied to execute a global sensitivity analysis and to analyze both the sensitivity of the spectral hemispherical reflectance to the defined input parameters and the relative contribution of each of the parameters to the model output.

To obtain the training data for the GP model we generated a 250 point Latin Hypercube design over the 7-dimensional space of inputs given in table I. The Leaf Angle Distribution (LAD) variable was set to planophile (leaves mostly horizontal) and the sun angle was set to zenith. While the sun angle will vary, for any given satellite scene it will be known, and so we do not consider it as one of the inputs for this analysis. The LAD will be determined largely by knowledge of the biome of

TABLE II
THE WAVELENGTH FOR EACH BAND USED, AND THE CORRESPONDING MODIS BAND NUMBER.

band number	wavelength (nm)	MODIS band
1	469	ref3
2	555	ref4
3	1240	ref5
4	1640	ref6
5	2130	ref7
6	667	ref13
7	748	ref15
8	870	ref16

the area being observed. The LCM was run at 8 wavelengths, given in table II, corresponding to eight of the MODIS bands that are sensitive to vegetation. The corresponding MODIS band is also given in table II. Note that the bands are in MODIS band order, not in wavelength order.

Figure 2 shows plots of the main effects for the 7 input variables for each of the 8 bands. The larger the variation of the main effect plot, the greater the influence of that input on the LCM response. To display the main effects for all parameters on a single plot, the range of each input (given in table I) has been normalized to 0-1. The slope of each main effect plot gives information as to whether the output is an increasing or decreasing function of that input. The relative scale of the main effects can be easily compared visually. The absolute scale depends on the absolute magnitude of the model output. Figure 3 shows the main effects for band 4 and includes the uncertainty bounds due to approximating the LCM by the GP. The uncertainties are extremely small.

To correctly interpret the results and to put them in the right perspective, we divide the input parameters in two categories, i.e., absorption and scattering driven. Biochemical inputs, i.e.,

chlorophyll, water, lignin and protein are absorption driven since their effect heavily depends on wavelength, and mainly affect the absorption characteristic of vegetation canopies [25]. Conversely, LAI, leaf thickness and soil brightness can be categorized as scattering driven since they directly influence the transport of photons in the medium. The main effects and sensitivity indices are analyzed next.

The LCM is most sensitive to LAI in the near-infrared (NIR) region of the spectrum (bands 7 and 8). It is shown in figure 2 that the LAI effect is highly non-linear and the behavior is such that in bands 7, 8 and 3 an increase in LAI produces an increase in reflectance. Conversely, the effect is opposite in the visible, i.e., increases in LAI produces a decrease in the hemispherical reflectance. This trend is known. [38], [39]. The sensitivity indices (table III) show that LAI is the major contributor in the bands which are most sensitive.

Chlorophyll, on the other end, is expected to be extremely influential in the visible – it is the prevailing factor that dominates the reflectance. Its effect is strong in the visible (bands 1, 2 and 6) while it dramatically decreases at the red-edge (band 7) to eventually disappear in the rest of the spectrum. The decrease of sensitivity to chlorophyll at the red edge is only found because the MODIS bands do not cover the actual red edge which is located around 730nm. Conversely, chlorophyll does not absorb light after 760nm. As shown in figure 3, band 4 shows basically no sensitivity for chlorophyll with small quantified uncertainty in the result.

Water contribution occurs mainly in the short-wave infrared since it exhibits higher absorption which peaks around 1445nm and 1950nm. Indeed, water is ranked as the second and third major contributor to the reflectance in bands 5 and 4, respectively. Nevertheless, such bands are outside the atmospheric window. Conversely the reflectance is weakly sensitive in the visible.

Protein is shown to be insensitive to most of the spectrum. Its effect as well as contribution are extremely small and can be only detected with difficulties in the NIR (e.g. band 8).

Lignin is one of the major surprising results. It is extremely sensitive in the short-wave infrared (bands 4 and 5) where it is also the major contributor to the hemispherical reflectance. This is mainly due to the strong absorption features in this part of the spectrum where lignin absorption coefficient features a peak around 2110nm.

Leaf thickness demonstrates a true scattering effect and its response shows interesting features. It is mainly sensitive and has the major contribution in bands 3, 7 and 8. We believe that what we are seeing is that changing the leaf thickness has more influence on scattering than on absorption. Specifically, as we change the leaf thickness the model assumes that the leaf mass is unchanged, meaning that the absorption has little effect as can be seen specifically in the NIR part of the spectrum.

The soil brightness has generally little effect. The spectrum for a typical soil was spectrally defined and the brightness parameter is responsible for increasing the soil hemispherical reflectance therefore simulating the dry-wet effect.

That the sensitivity indices do not sum to one indicates that interaction effects between two or more inputs are important in some bands, particularly bands 4 and 5. In future work we

TABLE III

THE SENSITIVITY INDICES FOR EACH INPUT FOR EACH SPECTRAL BAND.

input	band; wavelength (nm)							
	1 469	2 555	3 1240	4 1640	5 2130	6 667	7 748	8 870
LAI	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.56	0.08	0.00
Water	0.00	0.00	0.01	0.12	0.14	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Lignin	0.00	0.00	0.19	0.36	0.53	0.00	0.13	0.16
Thick.	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Soil	0.00	0.00	0.08	0.06	0.03	0.01	0.03	0.06
Total	0.88	0.90	0.86	0.80	0.81	0.87	0.90	0.90

will compute the second order sensitivity indices that quantify which interactions are important.

The results described so far are for planophile LAD. We performed the same analysis for the other LAD values, namely erectophile, extremophile and plagiophile. The results for the main effects were largely similar, with the main effect for LAI being less pronounced for erectophile compared to planophile in bands 4 and 5, and thickness showing less effect in band 7. That the effect of changing LAD is apparent in the scattering variables is to be expected.

Tables IV and V show the sensitivity indices for LAI and thickness for the four LAD values. It is seen that the thickness input has a larger effect for planophile LAD (especially in bands 7 and 8). Variation is seen for LAI when comparing erectophile LAD to the other values.

Our results are consistent with, and extend, previous statistically-based work. For example, [12] presented a methodology for sensitivity analysis based on design of numerical experiments aimed at providing a comparison between four canopy radiative transfer models coupled with a leaf-based radiative transfer model (PROSPECT, [29]). Their results are consistent with ours regarding LAI, chlorophyll and soil brightness sensitivity behavior. That the response in bands 1, 2 and 6 is dominated by LAI and chlorophyll is consistent with the results of a much more restricted sensitivity analysis in [40].

These results show that analyzing the uncertainty characteristics of RTMs used in remote sensed data product generation is practical and important. It gives information on the level of accuracy needed in the model's inputs, can guide data collection efforts to most effectively reduce the uncertainties, and can guide further development effort for the RTMs themselves. It also gives information as to which of the model's inputs affect the output, and hence which inputs it may be possible to determine from remotely-sensed observations.

VII. FUTURE WORK

In this paper we have developed a statistical framework for global sensitivity analysis in RTMs. In doing so we have introduced tools that have much wider applicability than the sensitivity analyses presented here. In particular, the GP emulator approach can be used to address important problems in model calibration, validation and inversion.

TABLE IV

THE SENSITIVITY INDICES FOR LAI FOR THE FOUR LAD VALUES.
(PLANOPHILE, ERECTOPHILE, EXTREMOPHILE AND PLAGIOPHILE)

LAD	band; wavelength (nm)							
	1	2	3	4	5	6	7	8
Plan.	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
Erect.	0.14	0.06	0.32	0.05	0.16	0.45	0.37	0.42
Extr.	0.08	0.03	0.40	0.12	0.09	0.36	0.40	0.47
Plag.	0.07	0.02	0.43	0.15	0.06	0.32	0.43	0.50

TABLE V

THE SENSITIVITY INDICES FOR THICKNESS FOR THE FOUR LAD VALUES.

Thick.	band; wavelength (nm)							
	1	2	3	4	5	6	7	8
Plan.	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Erect.	0.01	0.02	0.02	0.00	0.00	0.00	0.08	0.05
Extr.	0.02	0.04	0.08	0.02	0.02	0.01	0.18	0.12
Plag.	0.01	0.03	0.07	0.01	0.01	0.01	0.15	0.11

The remote sensing community spends much effort in collecting field data [41], [42] to calibrate and validate models – to determine how well a model matches reality, and to inform model improvements. Using the GP emulator approach, we can model the field data as

$$\tilde{y}_j = f(\tilde{\mathbf{v}}_j) + b(\tilde{\mathbf{v}}_j) + \epsilon_j \quad (10)$$

where \tilde{y}_j are the observed field data corresponding to parameter values $\tilde{\mathbf{v}}_j$, and the ϵ_j are measurement errors, say, independent from a $N(0, \sigma_\epsilon^2)$ distribution. The response is composed of two terms, $f(\cdot)$, the GP model based on training data, as described in section IV, and $b(\cdot)$, a bias term, a second GP which models the difference between the model approximation and the measured field data. (In practice, $f(\cdot)$ and $b(\cdot)$ are learned simultaneously based on a likelihood function that comprises both RTM training data and field data.) The resulting inference for $b(\cdot)$ for different regions of the input space can quantify the local performance of the model.

The determination of a calibrated, validated model incorporating the bias term allows a statistical inversion of the model to be performed, that respects the field data. In the usual manner, the error between the predictions and the satellite observations in a number of bands are taken to have a multivariate normal distribution. The inversion is regularized by the inclusion of a prior over LAI, which may depend on spatial position and biome type. Using the modeling framework in equation 10, the tools of statistical inference can be used to estimate the model inverse and its uncertainties.

The work outlined in this section is in progress, and will be reported when appropriate.

APPENDIX I

VARIANCE OF THE MAIN EFFECTS

We give here the derivation of $E^* \{ (E(Y | u_j))^2 \}$ required in the expression for $\text{Var}^* \{ E(Y | u_j) \}$, which provides a measure of the uncertainty associated with estimates of the main effects.

Note that,

$$\begin{aligned} (E(Y | u_j))^2 &= \left(\int_{\{v_\ell: \ell \neq j\}} f(v_1, \dots, u_j, \dots, v_d) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) \right)^2 \\ &= \iint_{\substack{\{v_\ell: \ell \neq j\} \\ \{v'_\ell: \ell \neq j\}}} f(v_1, \dots, u_j, \dots, v_d) f(v'_1, \dots, u_j, \dots, v'_d) \\ &\quad \times \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v'_\ell) \end{aligned}$$

and thus we need to take $E^* \{ \cdot \}$ with respect to the GP-based bivariate predictive distribution for $(f(v_1, \dots, u_j, \dots, v_d), f(v'_1, \dots, u_j, \dots, v'_d))$. Specifically,

$$\begin{aligned} E^* \{ (E(Y | u_j))^2 \} &= \iint_{\substack{\{v_\ell: \ell \neq j\} \\ \{v'_\ell: \ell \neq j\}}} E^* \{ f(v_1, \dots, u_j, \dots, v_d) f(v'_1, \dots, u_j, \dots, v'_d) \} \\ &\quad \times \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v'_\ell), \quad (11) \end{aligned}$$

where, using the standard covariance identity,

$$\begin{aligned} E^* \{ f(v_1, \dots, u_j, \dots, v_d) f(v'_1, \dots, u_j, \dots, v'_d) \} &= \\ \text{Cov}^* \{ f(v_1, \dots, u_j, \dots, v_d), f(v'_1, \dots, u_j, \dots, v'_d) \} &+ \\ + (E^* \{ f(v_1, \dots, u_j, \dots, v_d) \}) (E^* \{ f(v'_1, \dots, u_j, \dots, v'_d) \}). \quad (12) \end{aligned}$$

Denote by $\mathbf{R}_1 \equiv \mathbf{R}_1(v_1, \dots, u_j, \dots, v_d)$ and $\mathbf{R}_2 \equiv \mathbf{R}_2(v'_1, \dots, u_j, \dots, v'_d)$ the first and second columns, respectively, of the $(n \times 2)$ matrix $R(\mathbf{v}, \mathbf{v}')$ defined in section IV. Note that here the input vectors we are working with, $(v_1, \dots, u_j, \dots, v_d)$ and $(v'_1, \dots, u_j, \dots, v'_d)$, have common element u_j . Therefore, \mathbf{R}_1 is the $(n \times 1)$ vector with elements

$$\exp \left(-\frac{(u_j - x_{ij})^2}{\hat{\gamma}_j} - \sum_{\{\ell: \ell \neq j\}} \frac{(v_\ell - x_{i\ell})^2}{\hat{\gamma}_\ell} \right), \quad i = 1, \dots, n,$$

and analogously for \mathbf{R}_2 , replacing v_ℓ with v'_ℓ . Then, using (5) and (6), we obtain

$$\begin{aligned} E^* \{ f(v_1, \dots, u_j, \dots, v_d) \} &= \hat{\mu} + \mathbf{R}_1^T C^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n) \\ E^* \{ f(v'_1, \dots, u_j, \dots, v'_d) \} &= \hat{\mu} + \mathbf{R}_2^T C^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n) \\ \text{Cov}^* \{ f(v_1, \dots, u_j, \dots, v_d), f(v'_1, \dots, u_j, \dots, v'_d) \} &= \\ \hat{\sigma}^2 \left\{ \exp \left(-\sum_{\{\ell: \ell \neq j\}} \frac{(v_\ell - v'_\ell)^2}{\hat{\gamma}_\ell} \right) - \mathbf{R}_1^T C^{-1} \mathbf{R}_2 \right\}. \quad (13) \end{aligned}$$

Finally, substituting (12) and (13) in (11), we obtain for each $j = 1, \dots, d$,

$$\begin{aligned} E^* \{ (E(Y | u_j))^2 \} &= \hat{\sigma}^2 (e - \mathbf{T}_j^T(u_j) C^{-1} \mathbf{T}_j(u_j)) \\ &\quad + (\hat{\mu} + \mathbf{T}_j^T(u_j) C^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n))^2, \quad (14) \end{aligned}$$

where $\mathbf{T}_j(u_j)$ is the $(n \times 1)$ vector with elements given in equation 9 of section V, and

$$e = \prod_{\{\ell: \ell \neq j\}} \left\{ \int_{a_\ell}^{b_\ell} \int_{a_\ell}^{b_\ell} \exp\left(-\frac{(v_\ell - v'_\ell)^2}{\hat{\gamma}_\ell}\right) \frac{dv_\ell dv'_\ell}{(b_\ell - a_\ell)^2} \right\}. \quad (15)$$

Note that the second term in equation 14 is $(\mathbf{E}^*\{\mathbf{E}(Y | u_j)\})^2$, and so the required variance has the simpler expression

$$\text{Var}^*\{\mathbf{E}(Y | u_j)\} = \hat{\sigma}^2 (e - \mathbf{T}_j^T(u_j)C^{-1}\mathbf{T}_j(u_j)).$$

APPENDIX II THE SENSITIVITY INDICES

Here, we present the details for computing the estimates of the first-order sensitivity indices, $S_j = \text{Var}(\mathbf{E}(Y | u_j))/\text{Var}(Y)$, which provide a measure of the portion of variability in the response due to the main effect for each input. As discussed in section V, our estimates for the S_j , $j = 1, \dots, d$, are based on $\mathbf{E}^*\{\text{Var}(\mathbf{E}(Y | u_j))\}$ and $\mathbf{E}^*\{\text{Var}(Y)\}$. Regarding the estimate for the unconditional variance, we can write

$$\mathbf{E}^*\{\text{Var}(Y)\} = \mathbf{E}^*\{\mathbf{E}(Y^2)\} - \mathbf{E}^*\{(\mathbf{E}(Y))^2\}.$$

For the first term, we have

$$\begin{aligned} \mathbf{E}^*\{\mathbf{E}(Y^2)\} &= \mathbf{E}^*\left\{\int_{\mathbf{v}} f^2(\mathbf{v}) \prod_{\ell=1}^d dH_\ell(v_\ell)\right\} \\ &= \int_{\mathbf{v}} \mathbf{E}^*\{f^2(\mathbf{v})\} \prod_{\ell=1}^d dH_\ell(v_\ell) \\ &= \int_{\mathbf{v}} (S + m^2) \prod_{\ell=1}^d dH_\ell(v_\ell) \end{aligned}$$

where m and S are the mean and variance, respectively, of the predictive distribution for $f(\mathbf{v})$ given in section IV. Substituting their expressions to the above equation, we obtain

$$\begin{aligned} \mathbf{E}^*\{\mathbf{E}(Y^2)\} &= \hat{\sigma}^2 \int_{\mathbf{v}} (1 - \mathbf{r}^T(\mathbf{v})C^{-1}\mathbf{r}(\mathbf{v})) \prod_{\ell=1}^d dH_\ell(v_\ell) \\ &\quad + \int_{\mathbf{v}} \left\{ \hat{\mu}^2 + 2\hat{\mu}\mathbf{r}^T(\mathbf{v})C^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}_n) \right. \\ &\quad \left. + (\mathbf{r}^T(\mathbf{v})C^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}_n))^2 \right\} \prod_{\ell=1}^d dH_\ell(v_\ell) \\ &= \hat{\sigma}^2 - \hat{\sigma}^2 \left(\int_{\mathbf{v}} \mathbf{r}^T(\mathbf{v})C^{-1}\mathbf{r}(\mathbf{v}) \prod_{\ell=1}^d dH_\ell(v_\ell) \right) \\ &\quad + \hat{\mu}^2 + 2\hat{\mu}\mathbf{T}^T C^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}_n) \\ &\quad + \left(\int_{\mathbf{v}} (\mathbf{r}^T(\mathbf{v})C^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}_n))^2 \prod_{\ell=1}^d dH_\ell(v_\ell) \right) \end{aligned}$$

where \mathbf{T} is the $n \times 1$ vector defined in section V in the expression for $\mathbf{E}^*\{\mathbf{E}(Y)\}$ after equation 7. Regarding the two integrals above, if we expand the quadratic form $\mathbf{r}^T(\mathbf{v})C^{-1}\mathbf{r}(\mathbf{v})$ and apply the integral, we obtain

$$\int_{\mathbf{v}} \mathbf{r}^T(\mathbf{v})C^{-1}\mathbf{r}(\mathbf{v}) \prod_{\ell=1}^d dH_\ell(v_\ell) = \sum_{i=1}^n \sum_{j=1}^n c_{ij} q_{ij}$$

where c_{ij} is the (i, j) -th element of matrix C^{-1} , and

$$q_{ij} = \prod_{\ell=1}^d \left\{ \int_{a_\ell}^{b_\ell} \exp\left(-\frac{(v_\ell - x_{i\ell})^2 + (v_\ell - x_{j\ell})^2}{\hat{\gamma}_\ell}\right) \frac{1}{b_\ell - a_\ell} dv_\ell \right\},$$

$$i, j = 1, \dots, n.$$

(The q_{ij} are symmetric in (i, j) .) Analogously, expanding the square $(\mathbf{r}^T(\mathbf{v})C^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}_n))^2$ and taking the integral, we get

$$\begin{aligned} \int_{\mathbf{v}} (\mathbf{r}^T(\mathbf{v})C^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}_n))^2 \prod_{\ell=1}^d dH_\ell(v_\ell) \\ = \sum_{i=1}^n z_i^2 q_{ii} + 2 \sum_{i=1}^n \sum_{j=i+1}^n z_i z_j q_{ij} \end{aligned}$$

where z_i denotes the i -th element of the $n \times 1$ vector $C^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}_n)$.

For the second term, we can write

$$\begin{aligned} \mathbf{E}^*\{(\mathbf{E}(Y))^2\} &= \mathbf{E}^*\left\{\left(\int_{\mathbf{v}} f(\mathbf{v}) \prod_{\ell=1}^d dH_\ell(v_\ell)\right)^2\right\} \\ &= \mathbf{E}^*\left\{\int_{\mathbf{v}} \int_{\mathbf{v}'} f(\mathbf{v})f(\mathbf{v}') \prod_{\ell=1}^d dH_\ell(v_\ell) \prod_{\ell=1}^d dH_\ell(v'_\ell)\right\} \\ &= \int_{\mathbf{v}} \int_{\mathbf{v}'} \mathbf{E}^*\{f(\mathbf{v})f(\mathbf{v}')\} \prod_{\ell=1}^d dH_\ell(v_\ell) \prod_{\ell=1}^d dH_\ell(v'_\ell) \end{aligned}$$

and thus $\mathbf{E}^*\{(\mathbf{E}(Y))^2\}$ can be expressed as

$$\begin{aligned} \int_{\mathbf{v}} \int_{\mathbf{v}'} \text{Cov}^*\{f(\mathbf{v}), f(\mathbf{v}')\} \prod_{\ell=1}^d dH_\ell(v_\ell) \prod_{\ell=1}^d dH_\ell(v'_\ell) \\ + \int_{\mathbf{v}} \int_{\mathbf{v}'} \{\mathbf{E}^*(f(\mathbf{v}))\mathbf{E}^*(f(\mathbf{v}'))\} \prod_{\ell=1}^d dH_\ell(v_\ell) \prod_{\ell=1}^d dH_\ell(v'_\ell). \end{aligned} \quad (16)$$

Let $\mathbf{r}'(\mathbf{v}')$ denote the $(n \times 1)$ vector with i -th element given by $\exp(-\sum_{\ell=1}^d (v'_\ell - x_{i\ell})^2/\hat{\gamma}_\ell)$. Then, analogously to the expressions in (13), we have

$$\begin{aligned} \mathbf{E}^*\{f(\mathbf{v})\} &= \hat{\mu} + \mathbf{r}^T(\mathbf{v})C^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}_n) \\ \mathbf{E}^*\{f(\mathbf{v}')\} &= \hat{\mu} + \mathbf{r}'^T(\mathbf{v}')C^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}_n) \\ \text{Cov}^*\{f(\mathbf{v}), f(\mathbf{v}')\} &= \hat{\sigma}^2 \left\{ \exp\left(-\sum_{\ell=1}^d \frac{(v_\ell - v'_\ell)^2}{\hat{\gamma}_\ell}\right) \right. \\ &\quad \left. - \mathbf{r}^T(\mathbf{v})C^{-1}\mathbf{r}'(\mathbf{v}') \right\}. \end{aligned} \quad (17)$$

Therefore, substituting (17) in (16), and applying the integrations, we finally obtain

$$\begin{aligned} \mathbf{E}^*\{(\mathbf{E}(Y))^2\} &= \hat{\sigma}^2 (e^* - \mathbf{T}^T C^{-1} \mathbf{T}) + (\hat{\mu} + \mathbf{T}^T C^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}_n))^2, \end{aligned}$$

where

$$e^* = \prod_{\ell=1}^d \left\{ \int_{a_\ell}^{b_\ell} \int_{a_\ell}^{b_\ell} \exp\left(-\frac{(v_\ell - v'_\ell)^2}{\hat{\gamma}_\ell}\right) \frac{1}{(b_\ell - a_\ell)^2} dv_\ell dv'_\ell \right\},$$

and, again, \mathbf{T} is the $(n \times 1)$ vector given after equation 7 in section V.

Turning to the estimate for $\text{Var}(E(Y | u_j))$, we have

$$E^* \{ \text{Var}(E(Y | u_j)) \} = E^* \{ E [(E(Y | u_j))^2] \} - E^* \{ (E(Y))^2 \}$$

and therefore we only need the expression for $E^* \{ E [(E(Y | u_j))^2] \}$. In particular,

$$\begin{aligned} E^* \{ E [(E(Y | u_j))^2] \} &= E^* \left\{ \int (E(Y | u_j))^2 dH_j(u_j) \right\} \\ &= \int E^* \{ (E(Y | u_j))^2 \} dH_j(u_j) \\ &= \int \left\{ \hat{\sigma}^2 (e - \mathbf{T}_j^T(u_j) C^{-1} \mathbf{T}_j(u_j)) \right. \\ &\quad \left. + (\hat{\mu} + \mathbf{T}_j^T(u_j) C^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n))^2 \right\} dH_j(u_j) \\ &= \hat{\sigma}^2 e - \hat{\sigma}^2 \left\{ \int \mathbf{T}_j^T(u_j) C^{-1} \mathbf{T}_j(u_j) dH_j(u_j) \right\} \\ &\quad + \hat{\mu}^2 + 2\hat{\mu} \mathbf{T}^T C^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n) \\ &\quad + \left\{ \int (\mathbf{T}_j^T(u_j) C^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n))^2 dH_j(u_j) \right\}. \end{aligned}$$

The two integrals above can be computed as follows. First,

$$\int \mathbf{T}_j^T(u_j) C^{-1} \mathbf{T}_j(u_j) dH_j(u_j) = \sum_{m=1}^n \sum_{k=1}^n A_m A_k c_{mk} I_{km}$$

where

$$I_{km} = \int_{a_j}^{b_j} \exp \left(- \frac{(u_j - x_{mj})^2 + (u_j - x_{kj})^2}{\hat{\gamma}_j} \right) \frac{1}{b_j - a_j} du_j$$

and where, again, c_{mk} is the (m, k) -th element of matrix C^{-1} , and

$$A_m = \prod_{\{\ell: \ell \neq j\}} \left\{ \int_{a_\ell}^{b_\ell} \exp \left(- \frac{(v_\ell - x_{m\ell})^2}{\hat{\gamma}_\ell} \right) \frac{1}{b_\ell - a_\ell} dv_\ell \right\},$$

$m = 1, \dots, n.$

Moreover,

$$\begin{aligned} &\int (\mathbf{T}_j^T(u_j) C^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n))^2 dH_j(u_j) \\ &= \sum_{m=1}^n z_m^2 \left\{ A_m^2 \int_{a_j}^{b_j} \exp \left(- \frac{2(u_j - x_{mj})^2}{\hat{\gamma}_j} \right) \frac{1}{b_j - a_j} du_j \right\} \\ &\quad + 2 \sum_{m=1}^n \sum_{k=m+1}^n z_m z_k A_m A_k I_{km} \end{aligned}$$

where I_{km} is defined above and z_m denotes the m -th element of the vector $C^{-1}(\mathbf{y} - \hat{\mu} \mathbf{1}_n)$.

ACKNOWLEDGMENTS

This work was supported by the NASA AISR program, through grant numbers NNG06G177G and NNX07AV69G.

REFERENCES

- [1] A. Huete and H. Liu, "An error and sensitivity analysis of the atmospheric- and soil-correcting variants of the NDVI for the MODIS-EOS," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 897–905, July 1994.
- [2] C. Justice, E. Vermote, J. Townshend, R. Defries, D. Roy, D. Hall, V. Salomonson, J. Privette, G. Riggs, A. Strahler, W. Lucht, R. Myneni, Y. Knyazikhin, S. Running, R. Nemani, Z. Wan, A. Huete, W. van Leeuwen, R. Wolfe, L. Giglio, J.-P. Muller, P. Lewis, and M. Barnsley, "The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 4, pp. 1228–1249, July 1998.
- [3] W. Barnes, T. Pagano, and V. Salomonson, "Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 4, pp. 1088–1100, July 1998.
- [4] S. Hook, G. Vaughan, H. Tonooka, and S. Schladow, "Absolute radiometric in-flight validation of mid infrared and thermal infrared data from ASTER and MODIS on the Terra spacecraft using the Lake Tahoe CA/NV, USA, automated validation site," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1798–1807, 2007.
- [5] X. Xiong, N. Che, and W. Barnes, "Terra MODIS on-orbit spectral characterization and performance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2198–2206, August 2006.
- [6] J. Morisette, F. Baret, and S. Liang, "Special issue on global land product validation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 7, pp. 1695–1697, 2006.
- [7] W. Yang, B. Tan, D. Huang, M. Rautiainen, N. Shabanov, Y. Wang, J. Privette, K. Huemmrich, R. Fenshold, I. Sandholt, M. Weiss, D. Ahl, S. Glower, R. Nemani, Y. Knyazikhin, and R. Myneni, "MODIS leaf area index products: From validation to algorithm improvement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 7, pp. 1885–1989, 2006.
- [8] B. Ganapol, L. Johnson, C. Hlavka, D. Peterson, and B. Bond, "LCM2: A coupled leaf/canopy radiative transfer model," *Remote Sensing of the Environment*, vol. 70, pp. 153–166, 1999.
- [9] Y. Knyazikhin, J. Martonchik, R. Myneni, D. Diner, and S. Running, "Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data," *Journal of Geophysical Research*, vol. 103, pp. 32 257–32 276, 1998.
- [10] A. Saltelli, K. Chan, and E. Scott, *Sensitivity Analysis*. Chichester: John Wiley and Sons, 2000.
- [11] A. Saltelli, "Sensitivity analysis: Could better methods be used?" *Journal of Geophysical Research*, vol. 104, no. D3, pp. 3789–3793, February 1999.
- [12] C. Bacour, S. Jacquemoud, Y. Tourbier, M. Dechambre, and J.-P. Frangi, "Design and analysis of numerical experiments to compare four reflectance models," *Remote Sensing of the Environment*, vol. 79, pp. 72–83, 2002.
- [13] R. Brun, P. Reichert, and H. Kuensch, "Practical identifiability analysis of large environmental simulation models," *Water Resources Research*, vol. 37, no. 4, pp. 1015–1030, April 2001.
- [14] P. Bowyer, F. Danson, and N. Trodd, "Methods of sensitivity analysis in remote sensing: Implications for canopy reflectance model inversion," in *Geoscience and Remote Sensing Symposium*, vol. 6, July 2003, pp. 3839–3841.
- [15] D. Mackay, "Introduction to Gaussian process," *NATO ASI Series F Computer and Systems Sciences*, vol. 168, pp. 133–166, 1998.
- [16] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [17] R. Neal, "Regression and classification using Gaussian process priors," in *Bayesian Statistics 6*, J. Bernardo, J. Berger, A. Dawid, and A. Smith, Eds. Oxford University Press, 1998, pp. 475–501.
- [18] J. Sacks, W. Welch, W. Mitchell, and H. Wynn, "Design and analysis of computer experiments," *Statistical Science*, vol. 4, pp. 409–435, 1989.
- [19] C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker, "Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments," *Journal of the American Statistical Association*, vol. 86, pp. 953–963, 1991.
- [20] W. Welch, R. Buck, J. Sacks, H. Wynn, T. Mitchell, and M. Morris, "Screening, predicting and computer experiments," *Technometrics*, vol. 34, pp. 15–25, 1992.
- [21] T. Santner, B. Williams, and W. Notz, *The Design and Analysis of Computer Experiments*, ser. Springer Series in Statistics. Springer, 2003.
- [22] J. Oakley and A. O'Hagan, "Probabilistic sensitivity analysis of complex models: a Bayesian approach," *Journal of the Royal Statistical Society, Series B*, vol. 66, pp. 751–769, 2004.

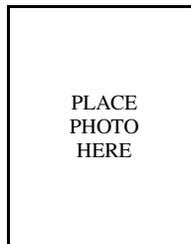
- [23] M. Kennedy and A. O'Hagan, "Bayesian calibration of computer models (with discussion)," *Journal of the Royal Statistical Society, series B*, vol. 63, pp. 425–464, 2001.
- [24] M. Bayarri, J. Berger, D. Higdon, M. Kennedy, A. Kottas, R. Paulo, J. Sacks, J. Cafeo, J. Cavendish, C. Lin, and J. Tu, "A framework for validation of computer models," in *Proceedings of the Workshop on Foundations for V and V in the 21st Century*, D. Pace and S. Stevenson, Eds. Society for Modeling and Simulation International, 2002, available as NISS Technical Report 2002-128.
- [25] B. Ganapol, L. Johnson, C. Hlavka, and D. Peterson, "LEAFMOD: A new within-leaf radiative transfer model," *Remote Sensing of the Environment*, vol. 63, pp. 182–193, 1998.
- [26] B. Hosgood, S. Jacquemoud, G. Andreoli, J. Verdebout, G. Peredrin, and G. Schmuck, "Leaf Optical Properties EXperiment (LOPEX93: Report EUR16095EN)," Joint Research Center-European Commission, Institute for Remote Sensing Applications, Tech. Rep., 1995.
- [27] P. Kubelka and F. Munk, "Ein Beitrag zur Optik der Farbanstriche," *Ann. Techn. Phys.*, vol. 11, pp. 593–601, 1931.
- [28] R. Clark and T. Roush, "Reflectance spectroscopy - quantitative analysis techniques for remote sensing applications," *Journal of Geophysical Research*, vol. 89, pp. 6329–6340, July 1984.
- [29] S. Jacquemoud and F. Baret, "PROSPECT: a model of leaf optical properties spectra," *Remote Sensing of the Environment*, vol. 34, no. 75-91, 1990.
- [30] B. Ganapol, "Vegetation canopy reflectance modeling with turbid medium radiative transfer," in *Computational Methods in Transport*, F. Graziani, Ed. Berlin: Springer, 2006, p. 173.
- [31] B. Ganapol and R. Myneni, "The FN method for the one-angle radiative transfer equation applied to plant canopies," *Remote Sensing of the Environment*, vol. 39, pp. 213–231, 1992.
- [32] W. Huang, Z. Niu, J. Wang, L. Liu, C. Zhao, and Q. Liu, "Identifying crop leaf angle distribution based on two-temporal and bidirectional canopy reflectance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 12, pp. 3601–3609, December 2006.
- [33] C. Bacour, F. Baret, D. Béal, M. Weiss, and K. Pavageau, "Neural network estimation of LAI, fAPAR, fCover and LAI \times C_{ab} from top of canopy MERIS reflectance data: Principles and validation," *Remote Sensing of the Environment*, vol. 105, pp. 313–325, 2006.
- [34] Z. Jiang, A. Huete, J. Li, and Y. Chen, "An analysis of angle-based with ratio-based vegetation indices," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2506–2513, September 2006.
- [35] R. Salvador, J. Piñol, S. Tarantola, and E. Pla, "Global sensitivity analysis and scale effects of a fire propagation model used over Mediterranean shrublands," *Ecological Modeling*, vol. 136, pp. 175–189, 2001.
- [36] M. McKay, R. Beckman, and W. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, pp. 239–245, 1979.
- [37] W. Gilks, S. Richardson, and D. Spiegelhalter, Eds., *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, 1995.
- [38] N. Goel, "Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data," *Remote Sensing Review*, vol. 4, pp. 1–222, 1988.
- [39] J. Clevers and W. Verhoef, "Modeling and synergistic use of optical and microwave remote sensing. LAI estimation from canopy reflectance and WDVI: a sensitivity analysis with the SAIL model," Netherlands Remote Sensing Board (BCRS), Tech. Rep. 90-39, 1991.
- [40] J. Dungan and B. Ganapol, "Sources of uncertainty in the prediction of LAI/PPAR from MODIS," in *AGU Fall Meeting*, San Francisco, 2002.
- [41] S. Running, D. Baldocchi, D. Turner, S. Gower, P. Bakwin, and K. Hibbard, "A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data," *Remote Sensing of Environment*, vol. 70, pp. 108–127, 1999.
- [42] W. Cohen, S. Gower, D. Turner, and S. Running, "Linking *in situ* measurements, remote sensing and models to validate MODIS products related to the terrestrial carbon cycle: boreal forests," in *LTER/GTOS Carbon Flux Scaling Workshop*, HJA Andrews Experimental Forest, Blue River, OR, May 2001, presentation available at <http://www.fsl.orst.edu/larse/bigfoot/pubs-pres.html>.



Robin Morris is a research scientist with the Universities Space Research Association. He is also Associate Adjunct Professor in the Department of Applied Mathematics and Statistics, University of California, Santa Cruz. He holds an MA in electrical and information sciences, and a PhD in signal processing, both from Cambridge University, UK. He was a postdoctoral researcher at INRIA, Sophia Antipolis, France, and at NASA Ames. His research interests are in Bayesian inference for scientific data analysis.



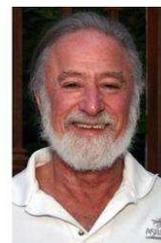
Athanasios Kottas is Associate Professor in the Department of Applied Mathematics and Statistics, University of California, Santa Cruz. He obtained his PhD in Statistics in 2000 from the University of Connecticut. His research interests include Bayesian nonparametric modeling for computer model experiments, regression problems, survival analysis, and spatial statistics, with applications in ecology and engineering.



Matthew Taddy is Assistant Professor of Econometrics and Statistics at the University of Chicago Graduate School of Business. He graduated B.A. (philosophy and mathematics) from McGill University, Montreal, (2003); M.S. (mathematical statistics) from McGill University, Montreal, (2005) and Ph.D. (statistics and stochastic modeling), University of California Santa Cruz, 2008.



Roberto Furfaro received a Laurea degree (master equivalent) in aeronautical engineering from Università di Roma La Sapienza in 1998 and a PhD in aerospace engineering from the University of Arizona in 2004. From 1998 to 2002 he held a position as researcher at the UA/NASA Space Engineering Research Center (SERC). From 2002 to 2004 he worked with the ecosystem science and technology branch at NASA Ames Research Center. Since 2005 he has been assistant research professor at the Aerospace and Mechanical Engineering Department, University of Arizona. He is an active member of the Global-Land and Ice Measurement from Space (GLIMS) consortium led by University of Arizona and member of the American Geophysical Union.



Barry Ganapol received a BS in mechanical engineering from University of California at Berkeley in 1966, a MS degree in engineering science from Columbia University in 1967 and a PhD in nuclear engineering from UC Berkeley in 1971. He is currently Professor in the Aerospace and Mechanical Engineering Department and in Hydrology and Water Resources, University of Arizona. Over the past 35 years he has been a visiting scientist at Los Alamos National Lab, NASA Ames, NASA Goddard, US Airforce Lab, Idaho National Lab, Oak Ridge National Lab and many universities around the world. In 2005 he received the G. Pomraning memorial award. In 2006 he was named Da Vinci Fellow at University of Arizona. He is a fellow of the American Nuclear Society.